

CONTACTLESS BIDIRECTIONAL NONVOLATILE MEMORY

Sau Ching Wong

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

[0001] This is a continuation-in-part of U.S. patent application No. 10/407,894, filed April 4, 2003, which is a divisional of U.S. Application No. 09/839,973, filed April 20, 2001, now U.S. Pat. No. 6,570,810.

BACKGROUND

[0002] A conventional contactless Flash memory array employs bit lines and source lines that are diffused into a silicon substrate. Fig. 1A conceptually illustrates the layout of part of a bank 100 in a conventional contactless Flash memory. In bank 100, n+ diffusion into a silicon substrate forms diffused bit lines 110 and diffused source (or ground) lines 120. Polysilicon floating gates 130 (poly 1) overlie channel regions, which are between diffused bit lines 110 and diffused source lines 120. Polysilicon word lines 140 (poly 2) cross over portions of diffused bit lines 110 and diffused source lines 120 that form the drains and sources of memory cells and also overlie associated floating gates 130.

[0003] Fig. 1B shows a cross section along a word line 140 in bank 100. As shown in Fig. 1B, channel regions 115 in the silicon substrate separate drain regions of diffused bit lines 110 from source regions of diffused source lines 120. Floating gates 130 overlie respective channel regions 115, with a gate insulator (e.g., gate oxide layer) between floating gates 130 and underlying channel regions 115. Word lines 140 overlie floating gates 130 with an insulating layer between each word line 140 and the underlying float gates 130 that are in a row corresponding to the word line.

[0004] Isolation structures 125 such as shallow trench isolation (STI) regions, conventional field oxide (LOCOS) regions, or heavily doped p+ field implant regions separate adjacent

diffused lines 110 from each other. Similar isolation structures (not shown) also separate adjacent channel regions 115 from each other.

[0005] As shown in Fig. 1A, each diffused bit line 110 extends to an associated bank select cell 170. Bank select cells 170 include transistors between respective diffused bit lines 110 and contacts to respective metal bit lines 180 that are typically part of a first metal layer and that overlies corresponding diffused bit lines 110. Generally, each metal bit line 180 extends over a number of banks and is connected to a corresponding select cell in each of the banks. A bank select line 160 controls bank select cells 170 in the bank to determine whether diffused bit lines 110 in the bank are connected to respective metal bit lines 180.

[0006] Diffused source lines 120 extend to contact virtual ground devices (not shown) or to other structures for control of the voltages of diffused source lines 120 during erase, write, and read operations.

[0007] A memory cell 150 in bank 100 of Fig. 1A includes a single floating gate transistor, and the number of memory cells in bank 100 is equal to the number of floating gates 130. In operation, charge added to or removed from the floating gate 130 in a memory cell 150 during a program or erase operation changes the threshold voltage of the floating gate transistor in the memory cell 150. The binary value stored in a memory cell 150 depends on whether the memory cell 150 is in a state having a high threshold voltage or a low threshold voltage.

[0008] The layout of each memory cell 150 in the contactless Flash memory array of Fig. 1A includes areas for a floating gate transistor and surrounding isolation structures. To minimize the area of a memory cell, features in the memory cells have widths or lengths equal to the minimum feature size, f , permitted by the design rules governing manufacture of the Flash memory integrated circuit. Along the direction perpendicular to diffused bit lines 110, the features for a memory cell include a shared isolation structure 125, a drain region of diffused bit line 110, a channel region 115, and a shared source region of diffused source line 120 as illustrated in Fig. 1B. The width and length of each of these structures are at least as wide or long as the minimum feature size f . Accordingly, a memory cell has an overall length (counting half the size of the shared features 120 and 125) of about $3f$. The overall width of the memory cell area includes the

width of the associated word line 140 and half of the widths of two adjacent isolation structures or at least $2f$. The minimum memory cell area is thus about $6f^2$ ($3fx2f$).

[0009] The effective memory cell size for bank 100 must additionally include a pro-rata portion of integrated circuit area associated with required overhead circuitry such as bank select cells 170, source contacts (not shown), and any virtual ground structures. More specifically, in one architecture M cells on a diffused bit line 110 share one bank select cell 170, the effective size of each memory cell is greater than $6f^2$ by $1/M$ times the area of a bank select cell and any other overhead for the column.

[0010] A contactless memory architecture that reduces the required cell area could increase the storage density achieved in a Flash memory integrated circuit (IC) and could therefore decrease the cost of Flash memory ICs.

SUMMARY

[0011] In accordance with an aspect of the invention, a contactless memory architecture has diffused lines that operate either as diffused bit lines or diffused source lines depending on the voltage through bank select cells to the diffused lines. Between each pair of diffused lines is a channel region, and overlying each channel regions is a charge trapping structure such as a conductive floating gate, a pair of conductive floating gates, or an insulator interface such as oxide-nitride interface. Word lines crossing the diffused lines overlie the channel and charge trapping structure between each pair of adjacent diffused lines, and no isolation structures are required between adjacent diffused lines. Elimination of these isolation structures significantly reduces the minimum cell area and permits a higher density of memory cells within a bank.

[0012] In the exemplary embodiment, each bank includes two sets of bank select cells, typically at opposite ends of the diffused lines. Each bank select cell except some corresponding to diffused lines at edges of a bank connect an associated metal line to a pair of diffused lines. One set of bank select cells connects the metal lines to pairs of diffused lines that are shifted relative to the pairs of diffused lines that the other set of bank select cells connect to the metal

lines. For any pair of adjacent diffused lines, activating one set of the bank select cells connects both of the adjacent diffused lines to the same metal line, and activating the other set of the bank select cells connects the two adjacent diffused lines to different metal lines.

[0013] The metal lines act either as bit lines or as ground lines depending on which column of memory cells an access operation selects and the desired direction of current through the selected memory cells. The layout of the metal lines across a set of banks generally includes a zigzag pattern to accommodate the relative shift of the bank select cells at opposite ends of each bank, but other layouts are possible.

[0014] An access to a memory cell in a selected column biases metal lines to one side (e.g., the left) of a selected column of memory cells at a first voltage and biases metal lines on the other side (e.g., to the right) of the selected column of memory cells at a second voltage. The appropriate set of bank select cells is activated for the access to achieve a voltage difference between the pair of adjacent diffused lines associated with the selected column. The resulting drain/source voltage difference of the memory cells in the selected column permits access of the selected memory cell in the selected column, and the common voltages on all other pairs of adjacent diffused lines prevent access or significant disturbance of memory cells in unselected columns. The first voltage can be higher or lower than the second voltage to select a direction for current through conducting memory cells in the selected column.

[0015] One embodiment of the invention is a bank of a memory such as a Flash memory. The bank includes memory cells arranged in rows and columns. Each memory cell includes a channel region in a substrate, a floating gate overlying the channel region, and a control gate overlying the floating gate. Diffused lines in the bank have portions that form source/drain regions of the memory cells, and every area of the substrate that is between an adjacent pair of the diffused lines contains the channel regions of memory cells that form a column in the bank.

[0016] The bank can additionally include first bank select cells and second bank select cells. Each first bank select cell includes a transistor between a corresponding adjacent pair of the diffused lines and a corresponding metal line. Each second bank select cell includes a transistor between a corresponding adjacent pair of the diffused lines and a corresponding one of the metal

lines, but the adjacent pairs of diffused lines corresponding to the second bank select cells are offset relative to the adjacent pairs of diffused lines corresponding to the first bank select cells. Additionally, one first bank select cell may be connected to a single diffused line, e.g., the first diffused line in the bank, and one second bank select cell may be connected to a single diffused line, e.g., the last diffused line in the bank.

[0017] Another embodiment of the invention is a memory including multiple banks and metal lines connected to the banks. Each bank includes: diffused lines in a substrate; channel regions arranged in rows and columns; floating gates respectively overlying the channel regions; word lines respectively overlying rows of the channel regions; first bank select cells; and second bank select cells. In each bank, each area of the substrate that is between an adjacent pair of the diffused lines contains a column of channel regions. The first bank select cells that are coupled to pairs of the diffused lines as are the second bank select cells, but the pairs of diffused lines coupled to the second bank select cells are offset relative to the pairs of diffused lines coupled to the first bank select cells.

[0018] Metal lines connect to respective first bank select cells and respective second bank select cells in each bank. Generally, the memory further includes column decoding circuitry coupled to bias the metal lines for an access of a selected memory cell. One embodiment of the invention includes thermometer-type column decoders in addition to conventional column decoders, drivers, sense amplifiers, and programming circuits. In response to a column address signal for the access, the thermometer-type column decoders bias all metal lines connected to diffused lines to a first side of the selected memory cell at a first voltage and biases all metal lines connected to diffused lines to a second side of the selected memory cell at a second voltage. Typically, the first voltage is ground, and the second voltage is positive and at a level required for a write or a read operation. The conventional column decoding circuits connect sense amplifier and programming circuits to the metal lines associated with the selected column.

[0019] In alternative embodiments of the invention, floating gates overlying the channel regions of the memory cells can be replaced with other charge trapping structures. One example of a charge trapping structure is an interface between insulating layers such as an interface

between a nitride layer and an oxide layer, and charge can be injected into and trapped in the interface during programming to change the threshold voltage of the memory cells. The trapped charge can be localized to two separated locations, where the location of charge injection depends on the current direction through the channel during programming. The separated charges cause the memory cell to be bidirectional in that the threshold voltage of the memory cell depends on the direction of the current through the channel. Accordingly, two distinct data values can be programmed and read from a single memory cell through control of the current direction during the programming and read operations.

[0020] Another suitable charge trapping structure that provides a bidirectional memory cell includes a pair of laterally spaced floating gates overlying the same channel. Current in one direction during programming injects charge into one of the floating gates, while current in the opposite direction injects charge into the other floating gates. The threshold voltage of the memory cell depends on the direction current or more particularly in an n-channel memory cell the amount of negative charge on the floating gate nearest the low voltage side of the memory cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Figs. 1A and 1B respectively illustrate a conceptual layout and a cross section of a conventional contactless Flash memory array with buried diffusion bit lines and source lines.

[0022] Fig. 2A shows a layout of a contactless Flash memory array in accordance with an embodiment of the invention.

[0023] Figs. 2B and 2C show cross sections respectively in column and row directions in the contactless Flash memory array of Fig. 2A.

[0024] Fig. 2D illustrates a layout for a bank selection cell for the contactless Flash memory array of Fig. 2A.

[0025] Figs. 3A and 3B illustrate operations accessing memory cells in selected columns of the contactless bank of Fig. 2A.

[0026] Fig. 4 is a timing diagram for signals employed during a programming operation in accordance with an embodiment of the invention.

[0027] Fig. 5 is a block diagram of a contactless flash memory in accordance with an embodiment of the invention.

[0028] Fig. 6A shows a layout of a contactless memory array in accordance with an embodiment of the invention including bidirectional memory cells.

[0029] Fig. 6B is a cross-section of the of the contactless memory array of Fig. 6A where each bidirectional memory cell includes two laterally spaced floating gates.

[0030] Fig. 6C is a cross-section of the of the contactless memory array in which each bidirectional memory cell includes a nitride layer have two laterally spaced areas that trap injected charge.

[0031] Use of the same reference symbols in different figures indicates similar or identical items.

DETAILED DESCRIPTION

[0032] In accordance with an aspect of the invention, a contactless memory architecture with buried diffusion bit/virtual ground lines reduces memory cell size by having memory cells between every pair of adjacent diffused lines and eliminating isolation structures that are between adjacent diffused bit lines in prior contactless Flash memories. In an exemplary embodiment of the invention, each diffused line can act as a diffused bit line or a diffused source line depending on which column of memory cells is selected. Two sets of bank select cells allow selection of either an odd or even column of memory cells for an access such as a write or read operation. Metal lines connect to the diffused lines via the bank select cells and bias the diffused lines as required for the access operation on the selected column.

[0033] Fig. 2A shows the layout of an M-by-N bank 200 of contactless Flash memory cells 250 in accordance with an embodiment of the invention. In a typical bank, the number of rows

M is 64, and the number of columns N could be up to 1024 or more for high-density memory. The bank includes N+1 diffused lines 210-0 to 210-N, generically referred to herein as diffused lines 210, and M word lines 240-1 to 240-M, generically referred to herein as word lines 240. Diffused lines 210 are in a p-well inside a deep n-well of an underlying p-type substrate and run in the direction of the N columns in bank 200. Word lines 240 are part of an overlying conductive layer (typically a second polysilicon layer) and run in the directions of the M rows of bank 200.

[0034] Portions of diffused lines 210 that are under word lines 240 form source/drain regions of floating gate transistors in respective memory cells 250. Each memory cell 250 further includes a channel region and a floating gate 230. The channel regions are in the p-well and between the associated source/drain regions of diffused lines 210, and floating gates 230 are between respective channel regions and associated word lines 240. Each floating gate 230 corresponds to a different memory cell 250 and is charged according to the value stored in the corresponding memory cell.

[0035] Fig. 2B shows a cross section of bank 200 that is parallel to and through a diffused line 210 formed in a semiconductor substrate. In the exemplary embodiment, the semiconductor substrate is a p-type substrate 202 containing a p-well 206 inside a deep n-well 204. Deep n-well 204 and p-well 206 have contacts (not shown) that permit independent biasing.

[0036] Diffused lines 210 are regions of n+ doping in p-well 206. Diffused lines 210 can be formed of salicide that is buried in the semiconductor substrate 202 to reduce the resistance of diffused lines 210. Formation of salicide for diffusion regions is well known for high-speed logic and memory processes, including contactless Flash memory arrays.

[0037] Fig. 2C shows a cross section of bank 200 that is parallel to and through a word line 240 and a diffused line 210. Diffused lines 210, which are formed before the second polysilicon (Poly2) layer during IC fabrication, cross under word lines 240, and portions of the diffused lines 210 that are under word lines 240 form source/drain regions of memory cells.

[0038] Returning to Fig. 2B, channel regions 215 are between diffused lines 210 and particularly between portions of diffused lines 210 that form n+ source/drain regions under word

lines 240. Unlike prior contactless Flash memories, no isolation structures are under word lines 240 to separate adjacent diffused lines. Isolation structures such as shallow trench isolation, field oxide isolation, or heavily doped p+ field implant regions (not shown) separate channel regions 215 that are in the same column of bank 200, i.e., between the same pair of diffused lines 210.

[0039] Floating gates 230 are typically from a first polysilicon layer, and a thin gate insulator such as a gate oxide layer separates the floating gates 230 from respective channel regions 215. Patterning of a second polysilicon layer forms word lines 240 that are over floating gates 230 with an insulating layer, typically an oxi-nitride-oxide (ONO) layer, between word lines 240 and the underlying floating gates 230.

[0040] Memory cells 250 in this embodiment of the invention are more compact than prior contactless Flash memories because adjacent memory cells 250 share both source/drain regions and isolation structures¹²⁵ (Fig. 1B) are eliminated. In particular, if the channel length of each channel 215 and the width of each diffused line is equal to the minimum feature size f, each memory cell 250 has a width of 2f, and the resulting memory cell area is $4f^2$ ($2fx2f$). The prior memory cells 150 as illustrated in Fig. 1B are thus 50% larger than memory cells 250.

[0041] As shown in Fig. 2A, N/2+1 metal lines 280-0 to 280-N/2 respectively connect to N/2+1 bank select cells 270-0 to 270-N/2 and N/2+1 bank select cells 275-0 to 275-N/2, which are generically referred to as bank select cells 270 and 275. Bank select cells 270 are at the ends of diffused lines 210 opposite to the ends connected to bank select cells 275. Excluding bank select cells 270-0 and 275-N/2, which are at edges of bank 200, each bank select cell 270 or 275 operates as a switch between the respective metal line 280 and a pair of adjacent diffused lines 210.

[0042] Connections of bank select cells 270 to respective diffused lines 210 are staggered relative to the connections of bank select cells 275 to respective diffused lines 210. More specifically, bank select cell 270-0 is between metal line 280-0 and diffused line 210-0. Bank select cell 275-0 is between metal line 280-0 and a pair of diffused lines 210-0 and 210-1. Bank select cell 270-1 is between metal line 280-1 and a pair of diffused lines 210-1 and 210-2, and bank select cell 275-1 is between metal line 280-1 and a pair of diffused lines 210-2 and 210-3.

This pattern continues up to bank select cell 270-N/2, which is between metal line 280-N/2 and diffused lines 210-(N-1) and 210-N, and bank select cell 275-N/2, which is between metal line 280-N/2 and diffused line 210-N.

[0043] Bank select lines 260 and 265, which can be formed from the layer (typically poly2) forming word lines 240, respectively control bank select cells 270 and 275. Activation of a select signal XA on bank select line 260 simultaneously turns on all bank select cells 270 in bank 200, so that bank select cells 270 electrically connect metal lines 280 to diffused lines 210. Activation of a bank select signal XB on bank select line 265 simultaneously turns on all bank select cells 275 in bank 200, and bank select cells 275 electrically connect metal lines 280 to diffused lines 210.

[0044] Fig. 2D shows an exemplary layout of a bank select cell 270. For faster response in charging of diffused lines 210, bank select cells 270 require a channel width that may be too large to be accommodated in the very small pitch of the columns of bank 200. Accordingly, gates 262 extend from bank select line 260 over channels of bank select cell 270. Each metal line 280 connects through one or more contacts to a drain region 272 in a central portion of the respective bank select cell 270. The channel regions under gates 262 separate each drain region 272 from associated source regions 273, and metal interconnects 274 contact source regions 273 and associated diffused lines 210.

[0045] Bank select cells 275 have a structure similar or identical to bank select cells 270. An adjacent bank (not shown) that is below bank 200 in Fig. 2A can be mirror-symmetric so that bank select cells 275 in the adjacent banks share contacts to metal lines 280. Similarly, an adjacent bank (not shown) that is above bank 200 in Fig. 2A can be mirror-symmetric so that bank select cells 270 in the adjacent banks share contacts to metal lines 280.

[0046] When accessing a memory cell in bank 200 of Fig. 2A, decoding circuitry (not shown) activates one of select signals XA and XB. In particular, when an address signal identifies a memory cell in bank 200, column decoding circuitry (not shown) activates select signal XA when the memory cell being access is in an odd-numbered column 1, 3, ..., N-1 and alternatively activates bank select signal XB when the memory cell being access is in an even-

numbered column 2, 4, ..., N.

[0047] Fig. 3A illustrates a write operation that uses channel hot electron injection to program a selected memory cell 322, which is in the second row and second column of bank 200. The timing for a programming operation is described further below, but during the programming operation, conventional row decoders and drivers (not shown) drive the selected word line 240-2 at a word line programming voltage Vpp, typically about 9 to 12 volts, depending on the target threshold voltage for the programming, the memory cell device characteristics, and the programming time budget. Unselected word lines 240 are biased low (e.g., grounded).

[0048] Bank select circuitry activates bank select signal XB on bank select line 265 to turn on bank select cells 275 in response to an address signal indicating that selected memory cell 322 is in an even column (column 2) of bank 200. As a result, bank select cells 275 connect metal lines 280-0 and 280-1 respectively to diffused lines 210-1 and 210-2, which are on opposite sides of selected memory cell 322. The select signal XA on bank select line 260 is deactivated to turn off bank select cells 270.

[0049] Column decoding and drive circuits (not shown) drive metal line 280-0 to a low voltage (typically ground) and drive metal line 280-1 to a programming voltage Vw (typically about 4.5 to 6 volts) depending on the target threshold voltage for the programming, the memory cell device characteristics, and the programming time budget. Since the selected word line 240-2 is at a high voltage and turns on selected memory cell 322, a current 310 flows from metal line 280-1 to metal line 280-0 through bank select cell 275-1, diffused line 210-2, the channel 215 of selected memory cell 322, diffused line 210-1, and bank select cell 275-0. Current 310 through channel region 215 of selected memory cell 322 causes channel hot electron injection into the floating gate 230 of selected memory cell 322 and raises the threshold voltage of selected memory cell 322.

[0050] During programming, bank select cell 275-1 also applies programming voltage Vw from metal line 280-1 to diffused lines 210-2 and 210-3. An unselected memory cell 323, which is in the selected row and between diffused lines 210-2 and 210-3, is not programmed because

both source/drain regions of unselected memory cell 323 are at substantially the same voltage and no significant current flows through unselected memory cell 323.

[0051] An unselected memory cell 324, which is in the selected row and between diffused lines 210-3 and 210-4, has a source/drain region portion in diffused line 210-3 at programming voltage Vw and a corresponding word line 240-2 at programming voltage Vpp. To prevent programming of unselected cell 324, column decoding and virtual ground/bit line (VG/BL) bias circuits raise metal line 280-2 to the programming voltage and thereby charges diffused line 210-4 via metal line 280-2 and bank select cell 275-2 to programming voltage Vw. Both source/drain regions of unselected cell 324 being at substantially the same voltage prevents programming of unselected cell 324.

[0052] The column decoding and VG/BL bias circuits prevent programming of all unselected memory cells that are in the selected row by biasing all metal lines and diffused lines on one side of the selected memory cell to the low voltage (virtual ground VG) and biasing all metal lines and diffused lines on the other side of selected memory cell to programming voltage Vw. In the particular example of Fig. 3A, column select and driver circuits ground metal line 280-0 and bias metal lines 280-1 to 280-N/2 to programming voltage Vw. With select signal XB activated, diffused lines 210-0 and 210-1 are at virtual ground VG, and diffused lines 210-2 to 210-N are at programming voltage Vw.

[0053] Fig. 3B illustrates a program operation when a selected memory cell 333 is in an odd column (column 3). For programming, the memory raises the selected word line 240-(M-1) to the word line programming voltage Vpp, grounds diffused line 210-2, and raises diffused line 210-3 to programming voltage Vw. To bias diffused lines 210, bank select logic activates the select signal XA on bank select line 260, and column decoders and VG/BL bias circuits ground metal lines 280-1 to 280-0 (i.e., metal lines to the left of selected cell 333) and bias metal lines 280-2 to 280-N/2 to programming voltage Vw. The bank select signal XB on bank select line 265 is deactivated to prevent connection of diffused lines 210-2 and 210-3 to the same metal line 280-1.

[0054] In the same manner as described above in regard to Fig. 3A, a current 315 flows from

metal line 280-2 to metal line 280-1 through bank select cell 270-2, diffused line 210-3, the channel 215 of the selected cell 333, diffused line 210-2, and bank select cell 270-1. Channel hot electron injection thus programs selected cell 333. Other memory cells in the selected row are not programmed because the unselected cells have substantially the same voltage (e.g., virtual ground VG or programming voltage Vw) on both source/drain regions.

[0055] For a write operation that does not program the selected memory cell, i.e., when the data input bit is 0 so that the cell is left in its “virgin” or unprogrammed state, the selected word line is biased at programming voltage Vpp, but all metal lines 280 are grounded or floating.

[0056] Table 1 illustrates the biasing of bank select lines 260 and 265 and metal lines 280 for an access of various columns of bank 200. As shown in Table 1, in the exemplary embodiment, metal line 280-0 is always grounded for an access operation, and metal line 280-N/2 is always at a positive voltage for the access operation.

Table 1: Bank Select and Metal Line Biasing for Programming (Data In = “1”)

	Selected column						
	1	2	3	4	...	N-1	N
XA (260)	High	Low	High	Low		High	Low
XB (265)	Low	High	Low	High		Low	High
280-0	Ground	Ground	Ground	Ground		Ground	Ground
280-1	Vw	Vw	Ground	Ground		Ground	Ground
280-2	Vw	Vw	Vw	Vw		Ground	Ground
280-3	Vw	Vw	Vw	Vw		Ground	Ground
...	Vw	Vw	Vw	Vw		Ground	Ground
280-N/2	Vw	Vw	Vw	Vw		Vw	Vw

[0057] The programming operations described above and illustrated in Figs. 3A and 3B can proceed in an order and with a timing that minimizes or prevents current through unselected memory cells while word line programming voltage Vpp is being applied. Fig. 4 illustrates the

timing of a programming operation that starts with all diffused lines 210, word lines 240, and metal lines 280 initially grounded. At a time T0, column decoders, drivers, and programming circuits drive the metal lines 280 that are to one side of the selected memory cell to programming voltage V_w . The metal lines 280 on the other side of the selected memory cell remain low at virtual ground VG. At the same time T0 or a slightly later time T1, bank select logic asserts the select signal XA or XB that corresponds to the bank and the column containing the selected memory cell. The activated bank select signal is at a voltage V_h that is greater than programming voltage V_w by an amount at least as large as the threshold voltages of bank select cells 270 and 275.

[0058] Row decoders and drivers delay activating the selected word line until a time T2. The delay allows current through the activated bank select cells 270 or 275 to charge diffused lines 210 to the desired voltages (virtual ground VG or programming voltage V_w) before the activation of the selected word line. This reduces transient currents that might otherwise flow through unselected memory cells in the selected row while diffused lines charge to the appropriate voltages. This minimizes the possibility of spurious programming of unselected memory cells on the selected word line. Similarly, at the end of the programming operation, the selected word line drops to ground at a time T3 before metal lines 280 and bank select line 260 or 265 return to ground at respective times T4 and T5.

[0059] Read operations use bank select cells 270 or 275 in a similar manner for selection of a memory cell in an odd or even column of bank 200 during a programming operation. In particular, a read operation activates select signal XA to turn on column select cells 270 when reading a memory cell in an odd column or activates select signal XB to turn on column select cells 275 when reading a memory cell in an even column.

[0060] During a read operation, row decoders and drivers bias the selected word line 240 to a read voltage V_{rr} that is greater than the threshold voltage of an unprogrammed memory cell but less than the threshold voltage of a programmed memory cell. The row decoders ground the unselected word lines 240. Biasing of bank select line 260 and 265 and metal lines 280 for the read operation is essentially the same as in the programming operation except that the positive

read voltages V_{rr} and V_r are lower than corresponding programming voltages V_{pp} and V_w .

[0061] Column decoders and VG/BL bias circuits ground the metal lines to one side of the selected column of memory cells and bias metal lines on the other side to a bit line read voltage V_r , typically less than 1 to 1.5 volts to avoid read disturb. The column decoder further connects a sense amplifier (not shown) to the selected metal line 280 that is at read voltage V_r and connected through a bank select cell 270 or 275 to the diffused line 210 that forms a drain region for the selected memory cell. Depending on its type, the sense amplifier senses current or a voltage drop of the metal line and generates an output signal indicating whether the selected memory cell is programmed or unprogrammed.

[0062] Only the selected memory cell is connected to the activated word line and has a voltage difference between its source/drain regions. Memory cells other than the selected memory cell that are connected to the activated word line have both source/drain regions at the same voltage (virtual ground VG or read voltage V_r) and do not conduct a current that would cause an error during the read operation. Accordingly, the presence or lack of current on the selected column or bit line depends solely on the threshold voltage state of the selected memory cell.

[0063] Typically, for an erase operation, a Flash memory erases one or more groups or sections of memory cells together, as a sector, by Fowler-Nordheim (F/N) tunneling of electrons out of the floating-gate through either the channel or source regions. A “negative-gate channel-erase” is preferred for the memory architecture of Figs. 2A, 2B, and 2C. For the negative-gate channel-erase method, the word lines 140 are biased to a negative erase voltage, typically -10 to -12 v, while P-well 206 and the deep N-well 204 are separately biased to 5-6 volts and at least 5-6 volts, respectively. The use of the negative-gate channel-erase eliminates the band-to-band tunneling current and associated ill effects that can arise with a “grounded-gate source-side erase”. Additionally, the “negative-gate channel-erase” does not require drain junctions to differ structurally from source junctions, which results in minimum cell size.

[0064] The size of p-well 206 limits the number of memory cells that are simultaneously erased as a sector. One or more memory banks 200 can reside in the same p-well 206 to permit

the simultaneous erasure of an entire bank or multiple banks of memory cells. To erase just selected rows of memory cells, one or more word lines 140 overlying the same p-well can be selectively biased to the negative erase voltage (-10 to -12 volts), while other unselected word lines 140 remain at near ground potential to avoid electrical erasure. With this architecture, a sector in memory 200 can be as small as a single row of a bank.

[0065] Fig. 5 illustrates a contactless Flash memory 500 in accordance with an embodiment of the invention. Contactless Flash memory 500 includes X banks that are similar or identical to bank 200. Metal lines 280 connect to each bank through contacts to bank select cells 270 and 275 in each bank. As shown in Fig. 5, the banks can be arranged in a mirror-symmetric fashion so that bank select cells 270 (or 275) for one bank shares a metal line contact structure with bank select cells 270 (or 275) in an adjacent bank.

[0066] For address decoding, contactless Flash memory 500 includes row decoder and bank selection circuits 510, column decoders 520, an even/odd decoder 530, and a thermometer-type column decoder 540.

[0067] Row decoding circuits 510 receive a row address signal, select a word line according to the row address signal, and drive the selected word line and unselected word lines as required for an access (write or read) of a selected memory cell. Additionally, row decoding circuits 510 activates one of the bank select signals B1 to BX according to which of the banks 1 to X contains the selected memory cell. Such row decoding circuits are well known for Flash memory and not described further here.

[0068] For activation of the appropriate select signal XA or XB, each bank includes a pair of AND gates 512 and 514, and each of the AND gate 512 and 514 has a corresponding one of bank select signals B1 to BX as an input signal. AND gates 512 have a signal O from even/odd decoder 530 as a second input signal, and AND gates 514 have a signal E from even/odd decoder 530 as a second input signal.

[0069] Even/odd decoder 530 activates signal O or E according to whether the selected memory cell is in an odd or even column of the array. In one embodiment, signal O is the least significant bit of the column address signal, and signal E has a level complementary to that of

signal O. Accordingly, one of AND gates 512 activates signal XA in the selected bank when the selected memory cell is in an odd column, and one of AND gates 514 activates select signal XB in the selected bank when the selected memory cell is in an even column.

[0070] Column decoders 520 and thermometer-type column decoders 540 are both connected to metal lines 280, and both decode the column address signal. Generally, column decoders 520 are conventional circuits that select a metal line 280 according to the column address signal. Column decoders 520 connect a programming circuit to the selected metal line 280 for a write operation or a sense amplifier to the selected metal line 280 for a read operation.

[0071] Thermometer-type decoders 540 connect VG/BL bias circuits to bias metal lines to one side of selected column at one voltage (e.g., virtual ground VG) and bias metal lines 280 on the other side of the select column at another voltage (e.g., programming voltage Vw or read voltage Vr). Circuits that perform thermometer-type decoding are known in the art and can be implemented in a variety of ways. Thermometer-type decoders 540 could be located next to or integrated into the conventional column decoders 520 or could be located at the ends of metal lines 280 opposite the conventional column decoder 520 as shown in Fig. 5.

[0072] For write and read operations, column decoder circuits 520 receive a column address signal and connect the selected metal lines 280 to a programming circuit or a sense amplifier. At the same time, thermometer column decoders selectively bias metal lines 280 on one side of (e.g., to the below) the selected column to virtual ground VG, and metal lines 280 on the other side of (e.g., above) the selected column to a positive voltage Vw or Vr. As described above, banks having N columns of memory cells have $1+N/2$ metal lines, and the column address not including the least significant bit identifies which of metal lines 280-1 to 280-N/2 is at the boundary between the metal lines biased at the positive voltage and the grounded metal lines.

[0073] As indicated above, embodiments of the invention provide compact memory cells for a contactless Flash memory by having memory cells between all adjacent pairs of diffused lines and using each diffused line as a virtual ground line or a bit line depending on which of the memory cells is being accessed. The effective size of memory cells including the overhead area required for bank select cells 270 and 275 can still be smaller than in conventional contactless

Flash memory arrays because bank select cells serve as both bank select cells and as virtual ground devices, while conventional contactless Flash memories often include both bank select cells and virtual ground devices.

[0074] The contactless memory architectures describe above can further be used with other types of memory cells. Fig. 6A is a layout view of a memory 600 containing bidirectional memory cells 650. Each bidirectional memory cell 650 is between corresponding pair of diffused lines 210 and includes a charge-storage or charge-trapping structure with two regions 631 and 632 that are laterally spaced apart over a common channel. With this structure, each bidirectional memory cell 650 corresponds to two data storage locations respectively associated with charge-storage or charge-trapping regions 631 and 632, and the two distinct binary or multi-level data values stored in the two storage locations correspond to the amount of charge respectively trapped in regions 631 and 632.

[0075] Memory 600 further includes word lines 240, bank select lines 260 and 265, bank select cells 270 and 275, and metal lines 280 that can be constructed and used as described above. In particular, an odd (or even) column of memory cells 650 can be selected by activating bank select signal XA (or XB) while applying a first voltage to all metal lines 280 to the left of the selected column and a second voltage to all metal lines to the right of the selected column. The separate data storage locations in each bidirectional memory cell 650 are independently accessed through control of whether the first voltage is higher or lower than the second voltage.

[0076] Fig. 6B shows a cross-sectional view along the line 6B-6B of an embodiment of memory 600 using a bidirectional memory cell 650 in which charge can be trapped at an interface between insulating materials. In the illustrated embodiment of Fig. 6B, memory 600 includes metal lines 280 and word lines 240 overlying a substrate 202 containing wells 204 and 206, diffused lines 210, and channel regions 215 as described above. Above each channel region 215 is a silicon nitride region 630 that is embedded in gate oxide. Programming of a memory cell 650 using channel hot electron injection will inject electrons into region 631 of silicon nitride 630 if the programming current is in one direction or will inject electrons into region 632 of silicon nitride 630 if the programming current is in the opposite direction. The insulating

properties of silicon nitride region 630 and the surrounding gate oxide traps the charge injected into region 631 in region 631 and traps the charge injected into region 632 in region 632. The threshold voltage that must be applied to word line 240 to make a particular memory cell 650 conduct depends on the direction of the channel current and the charge trapped in the region 631 or 632 nearest the source side memory cell 650. Accordingly, each memory cell 650 has two different threshold voltages, and each threshold voltage can be used to represent a binary or multiple-bit data value. U.S. Patent No. 6,011,725, entitled "Two Bit Non-Volatile Electrically Erasable and Programmable Semiconductor Memory Cell Utilizing Asymmetrical Charge Trapping," which is hereby incorporated by reference in its entirety, further describes the fabrication, properties, and uses of non-volatile memory cells of the type illustrated in Fig. 6B.

[0077] Fig. 6C shows a cross-sectional view of memory 600' for an embodiment that uses dual floating gates or electrically conductive regions 636 and 638 for charge storage regions 631 and 632. In the illustrated embodiment of Fig. 6C, memory 600' includes metal lines 280 and word lines 240 overlying a substrate 202 containing wells 204 and 206, diffused lines 210, and channel regions 215 as described above. Above each channel region 215 is a pair of floating gates 636 and 638 that can be made of a conductive material such as doped polysilicon. As described further in U.S. Pat. App. No. 10/140,527, entitled "Bi-Directional Floating Gate Nonvolatile Memory," which is hereby incorporated by reference in its entirety, programming of a memory cell 650' using channel hot electron injection will preferentially inject electrons into one region 636 or 638 depending on the direction of the programming current. Additionally, threshold voltage that must be applied to word line 240 to make a particular channel region 215 conduct depends on the direction of the channel current and the charge stored in the region 636 or 638 nearest the source line. Accordingly, each memory cell 650' has two different threshold voltages, and each threshold voltage can be used to represent a binary or multiple-bit data value.

[0078] Although the invention has been described with reference to particular embodiments, the description is only an example of the invention's application and should not be taken as a limitation. For example, although the above embodiments employ lines and structures made of particular materials such as metal or silicon or particular conductivity types such as n or p, other

materials or conductivity types are also suitable for use in embodiments of the invention. In particular, memory architectures in accordance with the invention can be applied in memories where the memory cells are p-channel devices. Additionally, although the above describes an embodiment of the invention where memory cells are a stacked-gate CHE/FN cells, other embodiments include array of other types of memory cells such as split-gate memory cells, source side-injection Flash memory cells, EPROM cells, nanocrystal memory cells, or even mask ROMs in which data values are according to a threshold voltage (V_t) implant. Such memories can store one bit per cell, multiple bits per cell, or analog values depending on the accuracy at which threshold voltages of memory cells can be written, retained, and read. Various other adaptations and combinations of features of the embodiments disclosed are within the scope of the invention as defined by the following claims.